Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: September 10, 2015

A. Przygienda
J. Tantsura
Ericsson

March 09, 2015

Automatic Assignment of BIER BFR-ids in ISIS
draft-prz-bier-bfrid-assignment-00

Abstract

Specification of an ISIS extension to support auto-election of BFR IDs in BIER using ISIS.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119] .

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 10, 2015.

Copyright Notice

Table of Contents

1. Introduction

Bit Index Explicit Replication (BIER) [I-D.draft-wijnands-bier-architecture-04] defines an architecture where all intended multicast receivers are encoded as bitmask in the Multicast packet header within different encapsulations such as [I-D.draft-wijnands-mpls-bier-encapsulation-02]. A router that receives such a packet will forward the packet based on the Bit Position in the packet header towards the receiver(s), following a precomputed tree for each of the bits in the packet. Each receiver is represented by a unique bit in the bitmask corresponding to its BFR-id. BFR-ids are sub-domain specific.

Once the number of receivers becomes large (i.e. many sets are present) or receivers choose to participate in many independent sub-domains, assignment of a unique BIER bit to a node is a non-trivial problem that can benefit highly from an automated solution. The

usual trade-offs are either a centralized (server) approach or a distributed approach which (from experience with other protocols such as DHCP or OSPF), provide at the cost of additional protocol complexity higher availability.

This document presents necessary, optional extensions to the currently deployed ISIS for IP [RFC1195] protocol to support automatic election of BFR-ids by means of a distributed protocol. This document defines new TLVs to be advertised by every router participating in BIER signaling and supporting such an election. In case some nodes are statically configured with a BFR-id, the protocol can detect misconfiguration, i.e. overlapping bit assignments or otherwise respects statically assigned BFR ids.

This extension operates seamlessly in a backwards compatible fashion with BIER procedures for ISIS as defined in [I-D.draft-przygienda-bier-isis-ranges-02]. Only BFRs implementing this extensions benefit from automatic assignment.

2. Terminology

Some of the terminology specified in [I-D.draft-wijnands-bier-architecture-04] is replicated here and extended by necessary definitions:

BIER: Bit Index Explicit Replication (The overall architecture of forwarding multicast using a Bit Position).

BIER-OL: BIER Overlay Signaling. (The method for the BFIR to learn about BFER's).

BFR: Bit Forwarding Router (A router that participates in Bit Index Multipoint Forwarding). A BFR is identified by a unique BFR-prefix in a BIER domain.

BFIR: Bit Forwarding Ingress Router (The ingress border router that inserts the BM into the packet).

BFER: Bit Forwarding Egress Router. A router that participates in Bit Index Forwarding as leaf. Each BFER must be a BFR. Each BFER must have a valid BFR-id assigned.

BFT: Bit Forwarding Tree used to reach all BFERs in a domain.

BIFT: Bit Index Forwarding Table.

BMS: Bit Mask Set. Set containing bit positions of all BFER participating in a set.

BMP: Bit Mask Position, a given bit in a BMS.

Invalid BMP: Unassigned Bit Mask Position, consisting of all 0s.

IGP signalled BIER domain: A BIER underlay where the BIER synchronization information is carried in IGP. Observe that a multi-topology is NOT a separate BIER domain in IGP.

BIER sub-domain: A further distinction within a BIER domain identified by its unique sub-domain identifier. A BIER sub-domain can support multiple BitString Lengths.

BFR-id: An optional, unique identifier for a BFR within a BIER sub-domain.

Invalid BFR-id: Unassigned BFR-id, consisting of all 0s.

3. IANA Considerations

This document adds the following new sub-sub-TLVs to the registry of sub-TLVs for BIER Info sub-TLV.

BIER Protocol Election sub-sub-TLV Value: TBD (suggested - to be assigned by IANA)

This document adds the following new TLV to the registery of ISIS TLVs.

BIER PE BMP Assignments TLV Value: TBD (suggested - to be assigned by IANA)

4. Procedures

The following sections present BIER IGP protocol procedures for the auto-election and maintainance of unique BIER BFR-ids across subdomains. Compared to purely administrative assignment of the bitmask use of those procedures largely facilitates deployment of BIER in large setups. The election and bit assignment procedures described in the according sections indicate how the BFRs participate in an election mechanism that allows them to

- use a dynamically chosen Designated and Backup Designated router for coordination and distribution of necessary state across all participants in the set across the network in a robust fashion
- allocate the necessary BMP in a sub-domain for each BFER

- automatically or administratively partition the elections for different sub-domains across the set of BFRs for maximum reliability
- discover administrative misconfiguration of BFERs

4.1. Election Algorithm

After a sub-domain <MT,SD,MLs> [I-D.draft-przygienda-bier-isis-ranges-02] is enabled, the according election procedures for D-BFR and Backup D-BFR are performed based upon the set of available BIER-PE sub-sub-TLVs. Given the fact that SD is uniquely tied to its MT per today's architecture and MLs are of no further importance to the introduced procedures, a sub-domain will be abbreviated without loss of generality as <SD>.

The election is indebted to and largely modeled (to the point of quoting parts of it verbatim) after the DR OSPF Election procedure in OSPF [RFC2328] which has proven to work exceedingly well over many years in the field.

This section describes the algorithm used for calculating a network's Designated BFR and Backup Designated BFR and procedures that allow those to allocate bit mask bits to a participating BFER in a sub-domain SD which we designate as BFER<SD>. The election runs per SD the router is participating in. The initial time a router runs the election algorithm, the D-BFR<SD> and BD-BFR<SD> are initialized to 0.0.0.0 or equivalent empty router ID. This indicates the lack of both a Designated BFR<SD> and a Backup Designated BFR<SD>.

The D-BFR<SD> election algorithm proceeds as follows:

- Call the router doing the calculation Router X<SD>. A router can participate in multiple elections for other BMS and multi-topologies at varying priorities.
- The list of BFRs participating in <SD> whose according BIER-PEs<SD> have been received by Router X<SD> and are connected (i.e. reachable via SPF computation) in standard topology MUST be examined.
- Router X<SD> itself MUST be also considered to be on the list.
- Discard all routers from the list that are ineligible to become D-BFR<SD>. (Routers having Router Priority of 0 for <SD> MUST NOT be eligible to become D-BFR<SD>.)

The following steps MUST then be executed, considering only those routers that remain on the list:

1. Note the current values for D-BFR<SD> and Backup D-BFR<SD>. This is used later for comparison purposes.

2. Calculate the new Backup D-BFR<SD> as follows.

   * Only those routers on the list that have not declared themselves to be D-BFR<SD> MUST be eligible to become Backup D-BFR<SD>.

   * If one or more of these routers have declared themselves Backup D-BFR<SD> (i.e., they are currently listing themselves as Backup D-BFR<SD>, but not as D-BFR<SD>, in their according BIER-PE packets) the one having highest Router Priority for <SD> MUST be declared to be Backup D-BFR<SD>.

   * In case of a tie, the one having the highest Router ID XOR'ed with SD (assuming big endian order, both values right-aligned and all bits of the shorter value filled up with zeroes to the length of the longer value) MUST be chosen.

   * If no routers have declared themselves Backup D-BFR<SD>, the router having highest Router Priority for <SD> MUST be chosen, (again excluding those routers who have declared themselves D-BFR<SD>), and again use the Router ID XOR'ed with SD to break ties.

3. Calculate the new D-BFR<SD> for the network as follows. If one or more of the routers have declared themselves D-BFR<SD> (i.e., they are currently listing themselves as D-BFR<SD> in their BIER-PE advertisements) the one having highest Router Priority for <SD> is declared to be D-BFR<SD>. In case of a tie, the one having the highest Router ID XOR'ed with SD is chosen. If no routers have declared themselves D-BFR<SD>, assign the D-BFR<SD> to be the same as the newly elected BD-BFR<SD>.

4. If Router X<SD> is now newly the D-BFR<SD> or newly the BD-BFR<SD>, or is now no longer the D-BFR<SD> or no longer the BD-BFR<SD>, repeat steps 2 and 3, and then proceed to step 5. For example, if Router X<SD> is now the D-BFR<SD>, when step 2 is repeated X<SD> will no longer be eligible for BD-BFR<SD> election. Among other things, this will ensure that no router will declare itself both BD-BFR<SD> and D-BFR<SD>.

5. As a result of these calculations, the router itself may now be D-BFR<SD> or BD-BFR<SD>. See Section 4.2 and Section 4.3 for the additional duties this would entail.

6. If the above calculations have caused the identity of either the D-BFR<SD> or BD-BFR<SD> to change, all routers must re-evaluate whether they have been elected D-BFR<SD> or BD-BFR<SD> and initiate according procedures. In case the new D-BFR<SD> or BD-BFR<SD> is not advertising according bitmask assignment and they are needed, they initiate according procedures in Section 4.2.1.

The reason behind the election algorithm's complexity is the desire for an orderly transition from BD-BFR<SD> to D-BFR<SD>, when the current D-BFR<SD> fails. This orderly transition is ensured through the introduction of hysteresis: no new BD-BFR<SD> can be chosen until the old Backup accepts its new D-BFR<SD> responsibilities.

The above procedure may elect the same router to be both D-BFR<SD> and BD-BFR<SD>, although that router will never be the calculating router (Router X<SD>) itself. The elected D-BFR<SD> may not be the router having the highest Router Priority for <SD>, nor will the BD-BFR<SD> necessarily have the second highest Router Priority. If Router X<SD> is not itself eligible to become D-BFR<SD>, it is possible that neither a BD-BFR<SD> nor a D-BFR<SD> will be selected in the above procedure. Note also that if Router X<SD> is the only router that is eligible to become D-BFR<SD>, it will select itself as D-BFR<SD> and there will be no BD-BFR<SD> for the network.

4.2. D-BFR Procedures

A router that assumes D-BFR role for a given <SD> combination invokes additional set of procedures as synchronization and election point for all the BFRs in <SD>.

4.2.1. Assignment of BMPs to BFERs in <SD>

Each BFER includes a strongly abbreviated DHCP-like FSM to obtain from the D-BFR<SD> its BMP or to advertise an administrative preference of its BMP.

The procedure is initiated by a BFER<SD> announcing in BIER Info sub-TLV for <SD> its assigned bit (or request for BMP assignment). The D-BFR<SD> initiates then a set of procedures to assign BMPs to such BFER in the <SD> or announces collisions.

Observe that BFERs can request (or announce) the bits even before a BDR<SD> has been chosen so the election and assignment are largely orthogonal sets of procedures.

4.3. BD-BFR Procedures

A router that is elected BD-BFR<SD> MUST mirror in its advertisements the exact state of the D-BFR<SD> and on each received advertisement maintains its internal states to use as starting point in all D-BFR<SD> procedures in case it looses connectivity (i.e. it cannot compute SPF reachability to the D-BFR in standard topology) to the D-BFR<SD>.

4.4. BFER Procedures

A BFER in <SD> controls its BMP in the set by providing values in its BIER Info sub-TLV for <SD> and signalling towards B-DR using A and R bits per Section 9.2. If it advertises the BFR-id without A or R bit set it indicates a fixed value it has chosen administratively.

It may request the assignment of a BMP by setting the R bit. The prefered BFR-id is signalled by providing a BFR-id value. The D-BFR MUST try to keep the preferred setting value when choosing BMP for the BFER. All other BFRs MUST NOT use the BFR-id value when the R bit is set. In case of routers not understanding this extensions, the behavior is enforced by the means of the C bit.

Once the BFER has been assigned a value from D-BFR and is willing to accept it, it MUST copy the value into the BFR-id field in the BIER-PE-BMPs it receives and set the A bit while clearing the R bit.

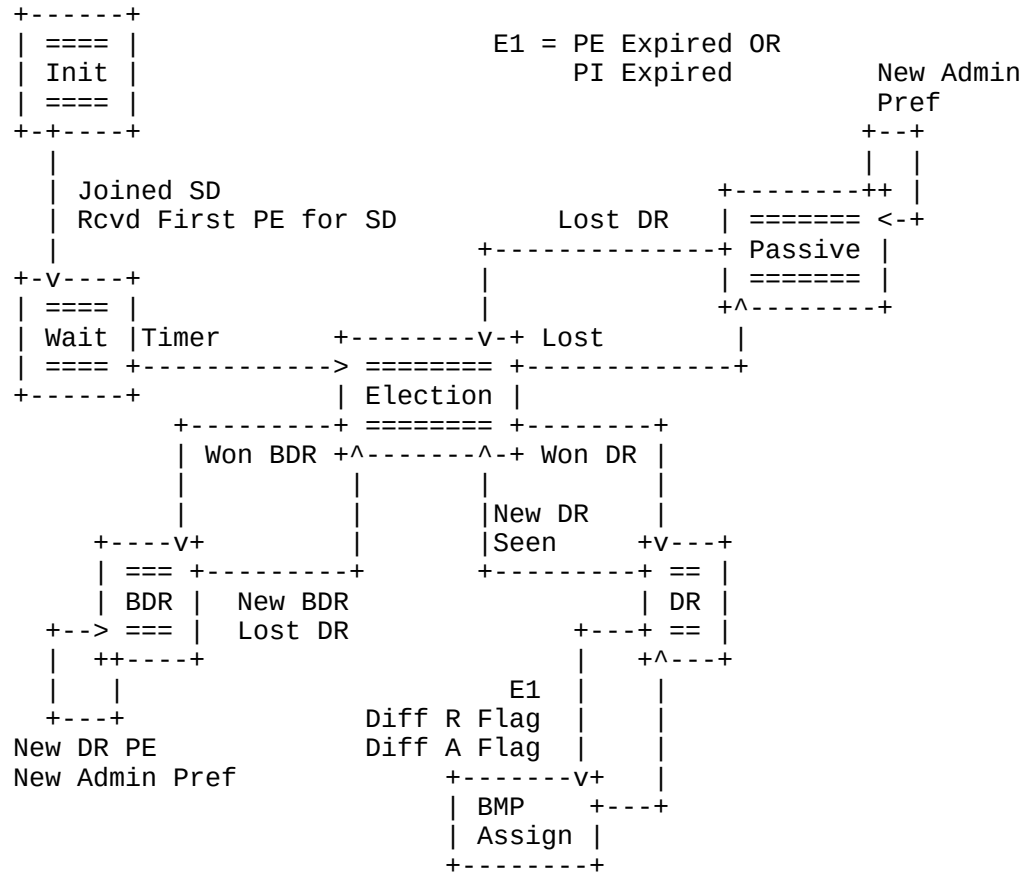
On the other side, the D-BFR for <SD> advertises the BMP assignments by the means of advertising BIER-PE-BMP for <SD>.

5. Special Considerations

5.1. BD-BFER to D-BFER Transition

In the normal case a router will assume its role as D-BFR<SD> promoting itself from BD-BFR<SD> with its own set of procedures. Based on those it will hold the state of the abdicating D-BFR<SD> and it MUST use this state as initial state for the D-BFR procedures it initiates per Section 4.2 . This should warranty a seamless fall-over without changes in the assignments of bits for BFERs for the according <SD> which SHOULD take preference over all other considerations. Observe that the implication is that a configured administrative preference MUST NOT be used unless changed or set explicitly again. The FSMs visualize this behavior more explicitly.

6. Election FSM for BFR<SD>



The full set of procedures can be described as a finite state machine per <SD> run within each participating BFR with the following events and transitions

6.1. States

Init Initial State of the Machine

Wait State waiting for routers to update their PEs for <SD> on startup

Election State that runs the election procedures and generates according events that progress it into another state immediately

Passive State entered when lost both DR and BDR in election.

Elected DR

Elected BDR

BMP Assign State in which the assignment of bits happens upons requests from BFERs.

6.2. Events

Timer Initial timer waiting for s of other routers before election is triggered.

Signalling/Rcvd First PE First PE for <SD> has been received or signalling enabled for the set S on BFR.

Lost DR Current D-BFR<SD> cannot be reached anymore via SPF computation in standard topology.

Lost Lost election for D-BFR and BD-BFR.

Won BDR Won election for BD-BFR.

Won DR Won election for D-BFR.

New BDR A new BD-BFR has been elected by the D-BFR.

New DR PE New BIER-PE Instance from D-BFR.

New Admin Pref Changed Administrative preference.

Diff R Flag R flag has been announced by a BFR which was not present before. In case of a new R flag, an assignment should be attempted. In case of R flag being deleted

if the A flag is set, the validity of the copied BFR-id with the assignment is checked

if the A flag is clear, the value is assumed non-negotiable and re-assignments may be necessary

Diff A Flag A flag has been withdrawn or announced. If A flag was present before and

R flag is clear, the value is assumed non-negotiable and re-assignments may be necessary.

R flag is set, a new assignment is requested.

If A flag was not present before and

R flag is clear, the validity of the copied BFR-id with the assignment is checked

R flag is set, the client MUST be declared faulty and disregarded.

To Be Completed TBD

7. FSM Figure/Events for BFER: TBD

8. Backwards Compatiblity

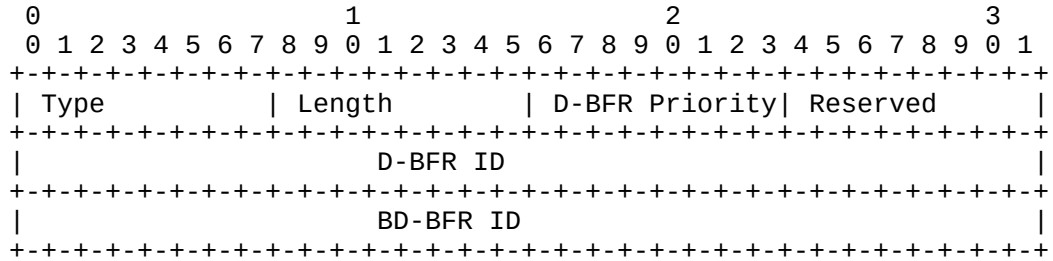
The procedures prescribed guarantee a complete backwards compatiblity to [I-D.draft-przygienda-bier-isis-ranges-02]. During the assignment procedure the according values are hidden from BFRs lacking this extension by the means of the C bit. Once assigned, they become visible. On the other hand, BFR-id values chosen by the BFRs without election extensions are respected in assignment.

9. Packet Formats

Some of the new information is carried within the the existing BIER Info sub-TLV per [I-D.draft-przygienda-bier-isis-ranges-02] and some presents a new ISIS TLV.

9.1. BIER-PE: BIER Protocol Election sub-sub-TLV

This sub-sub-TLV is included in the BIER Info sub-TLV of the according sub-domain as specified by [I-D.draft-przygienda-bier-isis-ranges-02]. It MUST be included in the BIER Info sub-TLV only once, otherwise the first instance is used.



Type: TBD1.

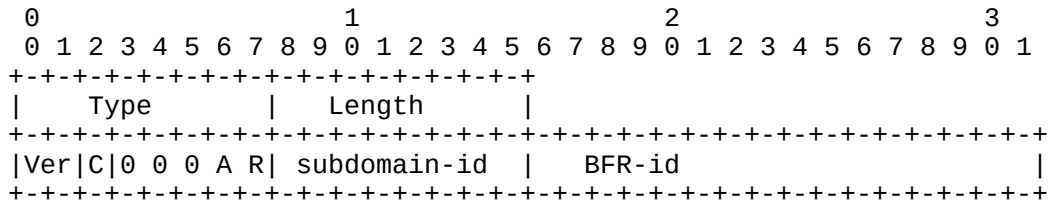
Length: 1 octet.

Priority Priority at which this router is set to become D-BFR for the <SD>.

D-BFR ID ID of the router chosen as D-BFR. If the router elected itself as D-BFR it MUST set it to its own ID.

BD-BFR ID ID of the router chosen as BD-BFR. If the router elected itself as BD-BFR it MUST set it to its own ID.

9.2. Reuse of the Reserved Bits in BIER Info sub-TLV



Version Version of the protocol. It remains at 0.

C The compatiblity bit. It is set according to following rules:

If the R bit is set, C is set to 0, i.e. the TLV is not compatible with version 0 of the BIER information. This will prevent routers not implementing this specification from looking at this advertisement.

If the R bit is clear, C is set to 1. In case the BFR-id has been obtained without an error by requesting it from a D-BFR, the value is copied into BFR-id of this sub-TLV, otherwise it is set to invalid BFR-id.

R Request Bit. When set, this bit advertises that the BFER is willing to accept another BMP than the one administratively

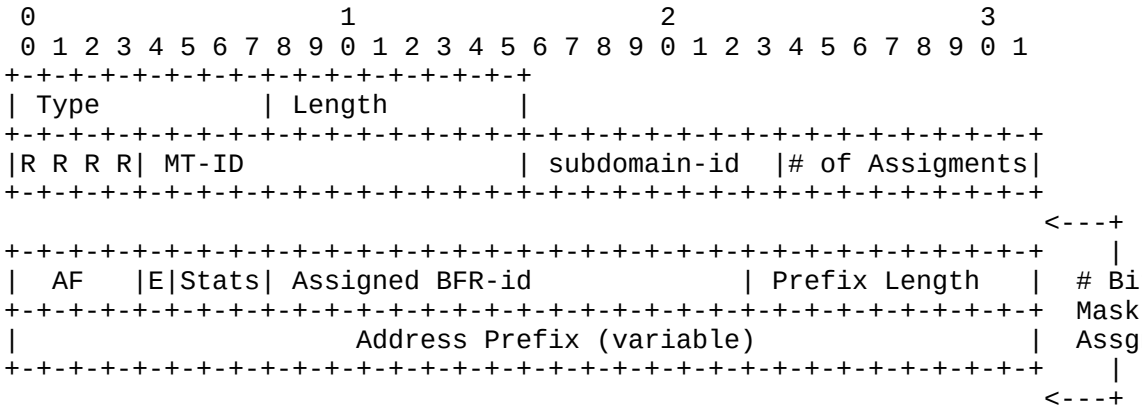
desired from D-BFR<SD>. The value of BMP is then determined by the according element in BIER-PE-BMP of the D-BFR<SD>.

A When this bit is set, the BFER advertises that the value indicated in the BFR-id has been copied from the assignment provided by D-BFR. If clear and BFR-id is set, the value is administratively assigned and is non-negotiable.

BFR-id If set and R bit is clear, it indicates the BFR-id the BFR is occupying to the D-BFR. If the R bit is set, it indicates the desired BFR-id to be assigned or no preference.

9.3. BIER-PE-BMP: BIER PE BMP Assignments TLV

This TLV is advertised only for the <SD> for which the router has been elected to be D-BDR<SD> or BD-BDR<SD>. It can repeat multiple times.



Type TBD

MT-ID Multi topology for which the assignments are provided

subdomain-id subdomain-id for which the assignments are provided

AF identifies address family of the prefix for which the assignment is provided. Values TBD

Prefix Length Prefix length of the prefix for which the assignment is provided.

Prefix Prefix containing the identifying prefix from TLVs 235, 237, 135 or 236 for which the assignment is provided.

Assigned BFR-id Bit Mask Position assigned by D-BFR, set to invalid BMP on an error status. 2 octets.

E Bit indicating assignment error, i.e. the BFER does NOT have a valid assignment.

Status Status of the assignment, 3 bits.

0 Assignment is OK and can be used (based on either administratively requested BMP or chosen by D-BFR for the requesting BFER). E-bit MUST be clear.

1 error: Unresolvable collision with other administratively set values, Bit Mask Position cannot be used. E-bit MUST be set.

2 error: Out of Bit Mask Positions for the Topology and Set, Bit Mask Position cannot be used. E-bit MUST be set.

all other values reserved, MUST NOT be used.

The assignments SHOULD be sorted on BFER-ID. Assignments MUST NOT repeat when the TLV is advertised multiple times and a router discovering such condition MUST issue an adequate warning. When multiple assignments for the same BFR are found, the first one in first TLV MUST be used and all others disregarded.

The assignments MUST NOT repeat any BIER Info sub-TLVs that have the R and A bit cleared, e.g. purely administrative assignments. A router discovering such condition MUST issue an adequate warning and disregard such assignments.

The assignments MUST repeat all assigned BIER Info sub-TLVs (that have A bit set). When such assignment is not advertised anymore, the according BFER MUST interpret that as loss as assignment, i.e. start with R bit again or set the BFR-id to invalid BFR-id.

10. Security Considerations

Implementations must assure that malformed TLV and sub-TLV permutations do not result in errors which cause hard protocol failures.

11. Acknowledgements

TBD.

12. Normative References

[I-D.draft-przygienda-bier-isis-ranges-02]
Przygienda et al., A., "BIER support via ISIS", internet-draft draft-przygienda-bier-isis-ranges-02.txt, Jan 2015.

[I-D.draft-wijnands-bier-architecture-04]
Wijnands, IJ., "Stateless Multicast using Bit Index Explicit Replication Architecture", internet-draft draft-wijnands-bier-architecture-04.txt, February 2015.

[I-D.draft-wijnands-mpls-bier-encapsulation-02]
Wijnands et al., IJ., "Bit Index Explicit Replication using MPLS encapsulation", internet-draft draft-wijnands-mpls-bier-encapsulation-02.txt, December 2014.

[RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, December 1990.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[RFC4971] Vasseur, JP., Shen, N., and R. Aggarwal, "Intermediate System to Intermediate System (IS-IS) Extensions for Advertising Router Information", RFC 4971, July 2007.

[RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, February 2008.

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.

[RFC5308] Hopps, C., "Routing IPv6 with IS-IS", RFC 5308, October 2008.

[RFC6513] Rosen, E. and R. Aggarwal, "Multicast in MPLS/BGP IP VPNs", RFC 6513, February 2012.

Authors' Addresses

Tony Przygienda
Ericsson
300 Holger Way
San Jose, CA 95134
USA

Email: antoni.przygienda@ericsson.com

Jeff Tantsura
Ericsson
300 Holger Way
San Jose, CA 95134
USA

Email: jeff.tantsura@ericsson.com